

영어 SentiWordNet을 이용하여 구축한 한국어 감성어휘사전의 성능 평가와 한계 연구

신동혁[○], 김새롬, 조동희, 뉘엔 민디오, 박순강, 어건주, 남지순
한국외국어대학교 언어인지과학과 DICORA

sdh876@hufs.ac.kr, steviasr@gmail.com, j.donghee@hufs.ac.kr, minkdieu@gmail.com, michellesg@hufs.ac.kr,
eoknj@hanmail.net, namjs@hufs.ac.kr

Performance and Limitations of a Korean Sentiment Lexicon Built on the English SentiWordNet

Donghyok Shin[○], Sairom Kim, Donghee Cho, Minh Dieu Nguyen, Soongang Park, Keonjoo Eo &
Jeesun Nam
DICORA Department of Linguistics and Cognitive Science, Hankuk University of Foreign Studies

요 약

본 연구는 다국어 감성사전 및 감성주석 코퍼스 구축 프로젝트인 MUSE 프로젝트의 일환으로 한국어 감성사전을 구축하기 위해 대표적인 영어 감성사전인 SentiWordNet을 이용하여 한국어 감성사전을 구축하는 방법의 의의와 한계점을 검토하는 것을 목적으로 한다. 우선 영어 SentiWordNet의 117,659개의 어휘 중에서 긍정/부정 0.5 스코어 이상의 어휘를 추출하여 구글 번역기를 이용해 자동 번역하는 작업을 실시하였다. 그 중에서 번역이 되지 않거나, 중복되는 경우를 제거하고, 언어학 전문가들의 수작업으로 분류해 낸 결과 3,665개의 감성어휘를 획득할 수 있었다. 그러나 이마저도 병명이나 순수 감성어휘로 보기 어려운 사례들이 상당수 포함되어 있어 실제 이를 코퍼스에 적용하여 감성어휘를 자동 판별했을 때에 맞집 코퍼스에서의 재현율(recall)이 긍정과 부정에서 각각 47.4%, 37.7%, IT 코퍼스에서 각각 55.2%, 32.4%에 불과하였다. 이와 더불어 F-measure의 경우, 맞집 코퍼스에서는 긍정과 부정의 값이 각각 62.3%, 38.5%였고, IT 코퍼스에서는 각각 65.5%, 44.6%의 낮은 수치를 보여주고 있어, SentiWordNet 기반의 감성사전은 감성사전으로서의 역할을 수행하기에 충분하지 않은 것으로 나타났다. 이를 통해 한국어 감성사전을 구축할 때에는 한국어의 언어적 속성을 고려한 체계적인 접근이 필요함을 역설하고, 현재 한국어 전자사전 DECO에 기반을 두어 보완 확장중인 SELEX 감성사전에 대해 소개한다.

주제어: 감성분석, 오피니언 마이닝, 셴티워드넷, 자동번역, 감성사전, 감성어휘, DECO한국어사전

1. 서론

본 연구¹⁾는 오피니언 마이닝(Opinion Mining) 시스템 개발을 위한 ‘MUSE²⁾ 감성³⁾사전 및 감성주석 코퍼스 구축 프로젝트’의 일환으로 수행되었다. 그중에서 영어권의 가장 대표적인 감성사전으로 평가되는 SentiWordNet을 이용하여 한국어 감성어휘사전을 구축할 때 나타나는 성과와 그 한계를 고찰하는 것을 목표로 한다.

2000년대 이후 트위터를 비롯한 SNS 사용으로 증가한

온라인 사용자 생성문(User-Generated Texts)의 주관적 의견을 분석할 수 있는 오피니언 마이닝(Opinion Mining)[1] 혹은 감성분석(Sentiment Analysis)[2]의 필요성이 대두되었다. 오피니언 마이닝 혹은 감성분석은 [3]에 따르면, '문어 텍스트에서 나타나는 개체나 속성에 대해 사람들의 의견, 감정, 평가, 태도, 감정을 분석하는 학문 영역'으로 정의된다. 최근 들어 오피니언 마이닝의 필요성이 증대되자, 이러한 시스템의 구축에 요구되는 감성사전의 필요성 또한 증가되었다.

그러나 사전을 구축하는 작업은 많은 시간과 비용이 요구된다. 이러한 현실적인 제약으로 인해 현재 오피니언 마이닝 분야에서 사용되는 감성사전은 일반적으로 일련의 알고리즘을 바탕으로 자동으로 구축된 사전을 사용하는 경향을 보이고 있다. 이들 연구는 주로 영어권을 중심으로 이루어져 있고, 영어권에서 이루어진 대표적인 감성사전의 성과로는 SentiWordNet[4], SenticNet[5], WordNet-Affect[6] 등이 있다. 이 중에서 가장 대표적으로 인용되고 많이 사용되고 있는 SentiWordNet(이후 SWN)을 이용하여 본 연구의 논의를 진행하고자 한다.

SWN을 자동번역기를 이용하여 다국어 연구에 활용하는

1) 본 연구는 2016년 미래창조과학부(한국연구재단) CRC 과제 일환으로 수행되었음(과제번호: 2015R1A5A7036384).
2) MUSE 프로젝트는 ‘Multilingual Sentiment Lexica and Sentiment-Annotated Corpora’를 구축하기 위한 연구로서 <http://dicora.hufs.ac.kr>를 참고할 것. 본 연구의 기초 데이터를 구축하고 검토하는 데에 한국외국어대학교 언어인지과학과 Arthur Provenier, 김소은, 한혜민, 류현영, 조다빈, 정선연, 김혜미 연구원이 함께 작업하였다. 이 자리를 빌어 감사의 뜻을 전한다.
3) 용어상의 문제에 있어, 본 연구에서는 국내 이미 통용되어 사용되고 있는 ‘감성’이라는 용어를 사용하기로 한다.

방법론에 대해서는 많은 비판이 있어왔다. 자동 번역의 문제를 언급한 [7]에서는 구글 번역기와 Bing, Moses라는 번역기를 이용했을 때, 39.2~61.2% 밖에 수행능력을 보여주지 못하고 있어 자동 번역을 신뢰할 수 없다는 주장을 하였다. 또한 SWN 자체에 대해서도 [8], [9]에서는 그 정확성의 한계에 대하여 문제를 제기하였다. 이렇듯 SWN에 대한 비판과 문제제기가 있어왔지만 SWN은 아직까지 감성사전으로서 가장 많이 인용되고, 실제로 가장 많이 사례연구에서 사용되고 있다.

본 연구에서는 영어 SWN을 이용하여 한국어 감성사전을 구축하는 방법의 의의와 한계점을 검토하고자 한다. 이를 통해 향후 SWN을 사용하고자 하는 연구들과 각 개별 언어의 감성사전을 구축하는 연구들에 하나의 사례연구를 제공할 수 있다는 점에서 기여한다고 생각한다.

본 연구의 각 장의 구성은 다음과 같다. 2장에서는 SWN을 이용하여 감성사전을 구축한 관련 연구를 간단히 소개하고, 3장에서는 SWN을 이용해서 한국어 감성사전을 구축하는 과정, 4장에서는 이렇게 구축된 감성사전을 검증하는 과정을 거쳐, 5장에서는 본 연구의 논의를 마무리할 것이다.

2. 관련 연구

SWN은 [4]에 의해 제안되었는데, 이 방법은 WordNet (Princeton WordNet, 이후 PWN)[10]의 어휘관계망을 이용하고 있다는 점에서 그 성능은 PWN에 상당부분 의존적이다. SWN은 14개의 기초 긍·부정 어휘를 이용하여, PWN의 관계망을 이용하여 자동적으로 극성의 스코어를 부착하는 방식을 이용하였다. 2016년 현재 3.0 버전[11]으로 117,659개의 어휘를 보유하며 영어권에서 가장 대표적인 감성사전으로 알려져 있다. 현재 SWN 홈페이지에서 파일을 다운로드할 수 있다(<http://sentiwordnet.isti.cnr.it/>).

SWN을 활용한 연구들은 현재까지도 계속해서 이어지고 있는데, 대부분 영어권 연구에 한정되어 있고, 이를 이용하여 다국어 감성사전을 구축한 사례는 실제로 그리 다양하지 않다. 이는 번역했을 때의 문제점을 어느 정도 인식하고 있기 때문으로 사료된다. 자동번역기를 거쳐서 이루어진 최초의 연구로는 [11]의 예를 들 수 있지만, 이것도 SWN을 번역한 것이 아니라, 독일어 문서를 영어로 번역한 뒤에 영어 SWN를 적용한 것이라 사실상 다국어 감성사전을 구축한 것은 아니라고 보아야 한다. 본격적으로 영어 SWN을 다른 언어로 번역하여 사용한 것은 [12], [13] 정도인데, 이들 연구는 SWN 어휘를 구글 번역기를 이용하여 번역한 결과, 인도 벵갈어, 힌디어, 텔루구어 어휘 각각 35,805개, 22,709개, 30,889개를 구축할 수 있었다고 보고되어 있다. 하지만 이중 최종 평가를 거친 언어는 벵갈어 하나였고, 나머지는 일부 평가에 거쳤다고만 언급하고 있다. 평가를 위해 사용된 벵갈어 뉴스와 블로그에서 72.16~74.6%의 정확률(precision)과 76~80.4%의 재현율(recall)로 높은 수치를 보여주었다고 보고하고 있다. 하지만 이는 SWN을 보완해주기 위한 일정 자질을 포함시켰을 때이고, SWN만을 사용했을 때에는 약 47.6%이라는 낮은 성능이 나타났다고 보고되어 있다.

한국어에서는 [8]에서 한국어 SWN을 구축하려는 시도가 있었으나, 제안 및 계획으로 제시되고 있어 구축된 결과를 실제로 확인할 방법은 없다. 이렇듯 SWN을 사용하여 다국어 감성사전을 구축하는 사례가 거의 확인되지 않음을 볼 수 있다. 이러한 상황에서 SWN을 한국어로 번역하여 감성사전을 구축하는 것은 향후 한국어 감성사전을 구축할 때, 어떠한 의의와 한계가 있을 것인지 이를 위한 사례연구로 참조할 수 있다는 점에서 본 연구의 의의를 찾을 수 있다.

3. SWN을 활용한 한국어 감성사전 표제어 구축

3.1. SWN 개요

앞서도 언급했지만, SWN은 PWN의 어휘관계망을 따라 긍·부정 및 객관의 스코어를 자동 부착하여 구축되었다. 여기서 나타난 감성사전의 수치는 합이 1이 되도록 정해져 있으며, 0~1 사이 범위에서 값이 존재하게 되어 있다. 다음 <표1>을 보자.

표1. SWN의 일부 예시

POS	Pos Score	Neg Score	Synset Terms	Gloss
a	0	0.625	bad#1	having undesirable or negative qualities; "a bad report card"
a	0	0.75	bad#12	characterized by wickedness or immorality; "led a very bad life"
a	0	0.375	bad#11	capable of harming; "bad air"; "smoking is bad for you"

위의 표는 현재 제공되는 SWN의 형식을 보여주고 있는데, SynsetTerms에서는 표제어가 나오고 다의어라는 표시가 #뒤에 숫자로 표시되어 나타난다. Gloss에서는 해당 표제어의 의미와 용례가 제시되고 있다. 이들에 대한 품사 정보는 왼쪽의 POS에 나타나 있는데 여기에서는 a(형용사)만 제시되어 있다. 각 표제어는 각각 PosScore(긍정 점수), NegScore(부정 점수)를 배당받는데, 객관은 1에서 이들을 뺀 값이다. 가령 bad#11의 경우, PosScore는 0, NegScore는 0.375이므로 1에서 이들을 뺀 값인 0.625가 객관 점수가 된다.

이렇듯 SWN은 각 어휘에 대해서 긍정 및 부정, 객관의 값을 제공하고 있기 때문에 잠재적 활용도가 높다고 할 수 있다. 그렇기 때문에 새로운 언어의 감성사전을 구축하고자 할 때, 초기 사전 구성에 유용한 기초 자료로 활용될 수 있다.

3.2. 구글 번역기를 이용한 SWN 감성어휘 추출

최대한 신뢰할 수 있는 감성어휘 목록을 구축하기 위해 현재 SWN에 수록된 117,659개의 어휘에서 긍정 및 부정 수치를 각각 0.5를 기준으로 하여 그 이상인 단어들을 추출하였다. 그 뒤에 표제어 형태만 고려해서 중복

제거를 실시하였고, 이렇게 획득된 감성어휘는 13,896개로 나타났다. 본래 내장된 단어 목록의 11.8%에 해당되는 비중이다. 중복 제거를 실시한 것은 현재의 구글 번역기가 한 단어에 대해 여러 한국어 대역어를 제공해주지 못하기 때문이었다. 예를 들면 'bad'를 번역기를 이용하면 '나쁜'만 실현되었다.

이렇게 획득된 13,896개의 영어 감성어휘를 구글 번역기를 이용해 한국어로 번역된 결과에서 다시 중복 제거하였고, 여기에 미번역 어휘를 제외하여 9,187개의 한국어 감성어휘 후보 목록을 산출하였다. 그런데 실제 번역된 결과물을 보면, 구글 번역기의 한국어 번역 수준이 매우 저조하다는 것을 확인할 수 있었다. 다음 <표2>는 한국어로 번역된 예시를 보여주고 있다.

표2. SWN의 한국어 번역 예

POS	Pos Score	Neg Score	English	Korean
n	0.625	0	going-over	가는 오버
n	0	0.5	family aloeaceae	가족 aloeaceae
v	0	0.5	steel onself for	강 onself을위한
v	0.875	0	feel like a million dollars	같은 백만 달러를 느낀다
n	0.625	0	white magic	병을 낮게 하거나 좋은 일이 생기게 하기 위하여 사용되는 선의의 마술
n	0	0.5	fowl cholera	닭 콜레라
n	0.5	0.125	glucosuria	당뇨

<표2>를 보면 '가는 오버', '같은 백만 달러를 느낀다'와 같이 무의미하거나 오역된 결과가 많이 나타났고, '가족 aloeaceae', '강 onself을위한'과 같이 중간에 미번역된 어휘들이 포함된 경우도 빈번하게 나타났다. 그리고 'white magic'처럼 대역어가 나타나지 않고, 설명식으로 번역되는 경우가 나타난 것을 볼 수 있는데, 이러한 상태에서 이들 번역 결과는 감성어휘 표제어로 사용되기에는 적합하지 않았다.

이밖에도 'family aloeaceae'와 같은 단어는 어떠한 종을 나타내는 단어임에도 부정 점수가 0.5로 나타나고 있는 것을 볼 수 있는데, 이러한 표제어를 감성어휘라고 할 수 있는지 의문스럽다. 이러한 예는 'fowl cholera'와 'glucosuria'에서도 나타나는데, 둘 다 병명임에도 전자는 부정, 후자는 긍정으로 스코어가 할당된 것은 적합하지 않아 보인다. 이러한 사례들은 SWN에서 상당수 나타나고 있다. 이러한 이유로 이들 어휘에 대해서 극성을 수작업으로 검토할 필요가 제기되었으며, 본 연구에서는 다음과 같은 방법을 이용하여 사전 구축을 진행하였다.

3.3. SWN 한국어 대역어에 대한 언어학적 검토

3.2에서 언급된 방법으로 추출된 9,187개의 어휘를 대상으로 세 사람의 언어학 전공자들의 중복 검토를 통해 작업의 신뢰도를 확보하였다. 번역된 한국어 대역어의 품사와 극성이 변하는 경우가 많았기 때문에, 이들의 극성과 품사를 다시 설정할 필요가 있었다. 그래서 SWN의 극성과 품사 정보를 모두 배제하고 [15], [17]의 DECO 한국어전자사전 극성 클래스와 품사 정보를 적용하였다. <그림1>은 극성 분류 정보를, <그림2>는 품사 분류 정보를 보인다.

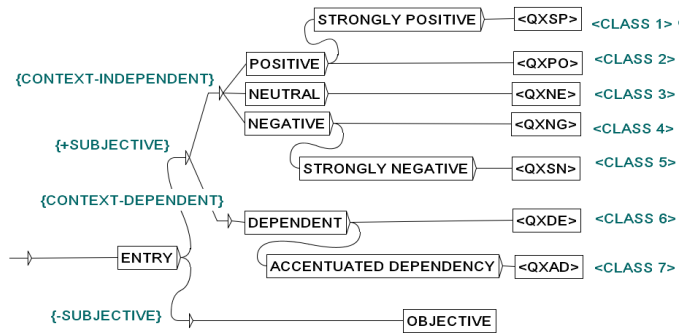


그림1. 한국어 감성어휘의 7가지 극성 정보

<그림1>의 {+Subjective/+Context-Independent}에서는 [15]에서 제안된 극성 클래스로 강함극성(QXSP)/긍정(QXPO)/중립(QXNE)/부정(QXNG)/강함부정(QXSN)의 5가지 강도로 우선 분류되고 {+2, +1, 0, -1, -2}의 5가지 점수로 세분화한 분류 방식을 보인다. {+Subjective/+Context-Dependent}에는 문맥에 의존적인 유형으로 상대극성(QXDE)/강한 상대극성(QXAD)의 2가지 유형이 분류되는데, 이러한 경우는 '많다' 같은 단어가 그 예가 될 수 있다. 가령 '많다'는 '인기가 많다'와 같은 문맥에서는 긍정, '먼지가 많다'와 같은 문맥에서는 부정 의견을 나타내는데 사용되었다. 이들은 문맥에 따라 결정되기 때문에 상대극성 어휘로 분류된다[16]. 마지막으로 주관적인 오피니언이 아닌 객관적인 어휘유형은 <OBJECTIVE>로 분류된다.

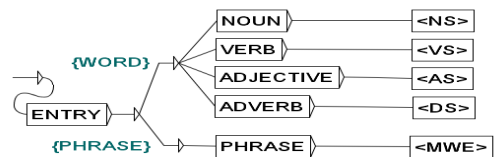


그림2. 한국어 감성어휘의 5가지 품사 정보

<그림2>는 [16]의 분류 체계에 의해 NS(명사 부류), AS(형용사 부류), VS(동사 부류), DS(부사 부류) 및 여러 단어 및 구 형태로 번역된 표제어는 MWE(Multi-Word Expressions: Phrase)라는 범주로 표시하였다.

이상과 같은 방법으로 7가지 극성 클래스로 분류된 3,665개(명사 2,409개/동사 409개/형용사 727개/부사 120개)의 감성어휘 표제어를 획득하였다. 다음 <표3>과 같다.

XML 방식의 마크를 부착하도록 LGG를 구축하여 획득한 결과를 보인다.

표3. SWN을 기반으로 구축된 감성사전 규모

TAG	NOUN	ADJECTIVE	VERB	ADVERB	TOTAL
QXSP	31	13	6	1	51
QXP0	650	275	104	52	1,081
QXNE	22	8	2	0	32
QXNG	1,425	339	221	47	2,032
QXSN	50	11	5	0	66
QXDE	222	79	70	19	390
QXAD	9	2	1	1	13
Total	2,409	727	409	120	3,665

(S) 갤럭시2로도 아직도 <positive>생생하</positive>게 잘쓰는 사람들 많음.(S)
 (S) 사이즈가 <positive>아담해</positive>서 오히려 여자들이 차면 더 멋있을듯.(S)
 † 공개됐을 때 그냥 좀 기다렸다가 옛지한번사볼까 했는데 삼성은 첫번째 폰은 항상 지달다는걸 깨달아.
 (S) 블랙이 제일 <positive>예쁜</positive>데 안파는게 이상하네요.(S)
 (S) 아이폰6는 <positive>완전한</positive> 혁신은 없지만 눈에 안보이는 혁신이 있는폰이다.(S)
 (S) 빨리 <positive>완전한</positive> 플렉서블 디스플레이를 채택한 스마트폰이나 나왔으면.
 (S) 전반적으로는 다 <positive>우세한</positive>데.(S)
 (S) 해상도가 4배 차이인데 노트5가 더 <positive>우월하</positive>거나 같기가 힘들.(S)
 어려울드리면 멀티태스킹 안되는 아이폰이 당연히 <positive>유리하</positive>지 8개앱이 같이들어가
 (S) 아이폰에 지프로배경화면있으니까 <positive>이쁘</positive>다.(S)
 (S) 요즘에는 <positive>이쁘</positive>다고 서로 막 사고 그러네요 ㅋㅋㅋ.(S)
 (S) 저도한달반정도 쓰는데 <positive>이쁘</positive>고 좋아요.(S)
 (S) V10 너무 <positive>이쁘</positive>길래 바로 V10샀는데 난 좋은데 이상없고.(S)

이렇게 구축된 어휘는 같은 구글 번역기를 이용한 [13], [14]의 벵갈어 3만 어휘와 비교할 때, 그 규모가 매우 작지만 개별 어휘에 대한 언어학적 검증이 수반되었다는 점에서 분명한 차별성을 가진다. 또한 다른 감성사전인 [5]의 SenticNet도 초기 5천 어휘로 시작된 점을 고려하면 초기 감성어휘라고 한정해 보았을 때, 3,665개라는 규모는 지나치게 작다고는 보이지 않는다. 다만 여기 구축된 감성사전에서는 감성어휘의 대부분을 차지하는 동사와 형용사의 수가 상대적으로 적고, 명사가 65.7%를 차지하였는데, 대다수의 명사가 '가와사키병, 각막염' 등의 병명으로 나타났다는 점이 문제라고 판단된다. 병은 일반적으로 부정적 함의를 가지고 있지만, 병명 전문용어 명사가 핵심적 감성어휘 부류를 구성하지 않기 때문이다. 이러한 점에서 SWN에만 기반을 두고 이를 번역하여 감성사전을 구축하는 것은, 비록 언어학적 검증을 거친다고 하여도, 그 표제어의 신뢰도에 적지 않은 문제점이 발생할 것이라 예상된다.

그림3. IT 리뷰 코퍼스의 긍정 어휘 추출 예시

여기에서는 형용사만을 대상으로 하여 검증을 실시하였는데, 실제적으로 형용사가 감성을 나타내는 어휘가 가장 많은 품사 부류이기 때문이다. 이렇게 코퍼스를 대상으로 SWN 기반으로 구축한 감성사전과 DECO 사전과의 비교를 통해서 정확률(precision)과 재현율(recall), 그리고 F-measure를 이용한 결과 다음과 같은 결과를 볼 수 있었다.

우선 맛집 리뷰 코퍼스를 대상으로 검증을 실시한 결과 1,200 문장에서 객관 문장 454 문장을 제외한 746 문장에 대해서, 정확률, 재현율, F-measure를 측정한 결과는 다음 <표4>와 같다. 그리고 IT 리뷰 코퍼스를 대상으로 검증을 실시한 결과는 1,200 문장에서 객관 문장 347 문장을 제외한 853 문장이 대상이 되었는데, 이때 정확률, 재현율, F-measure를 측정한 결과는 <표5>와 같다.

4. 구축된 한국어 감성사전의 검증

여기에서는 현재 구축된 SWN 기반 감성사전의 성능을 검증하기 위해서, 온라인상의 실제 사용자 리뷰 코퍼스를 대상으로 검증을 실시하였다. 사용된 코퍼스는 한국외국어대학교 DICORA 연구센터⁴⁾에서 구축된 사용자 IT 리뷰 {PECS-IT-100}와 맛집 리뷰 코퍼스 {PECS-RS-100}에서 각각 1200문장을 임의로 추출한 것으로, 여기서 관찰되는 실제 극성 어휘를 수작업으로 마크업해서 이를 평가셋으로 활용하였다. 대조군으로는 [17]의 한국어 DECO 사전을 이용하였다.

검증을 위해 UNITEX 3.1[18] 프로그램을 사용하였고, 유한 트랜스듀서 방식의 LGG(Local-Grammar Graphs)⁵⁾ 모델을 이용하여 추출하였다. <그림3>은 IT 리뷰 코퍼스에서 긍정 극성 어휘를 추출한 예를 보여주고 있는데, 긍정 어휘에 대해서 <positive>...</positive>와 같은

표4. 맛집 리뷰 코퍼스를 통한 두 사전의 재현율, 정확률, F-measure 비교

사전	극성	Recall	Precision	F-measure
SWN 기반	긍정	47.4	91.1	62.3
	부정	37.7	39.4	38.5
DECO 기반	긍정	69.1	87.2	77.1
	부정	47.8	66.0	55.4

표5. IT 리뷰 코퍼스를 통한 두 사전의 재현율, 정확률, F-measure 비교

사전	극성	Recall	Precision	F-measure
SWN 기반	긍정	55.2	80.3	65.5
	부정	32.4	71.0	44.6
DECO 기반	긍정	82.3	87.1	84.7
	부정	67.7	73.6	70.6

맛집 리뷰 코퍼스와 IT 리뷰 코퍼스를 대상으로 한 결과를 살펴보았을 때, SWN 기반 감성사전의 재현율은 전

4) <http://dicora.hufs.ac.kr>

5) LGG 모델은 유한상태 트랜스듀서(FST) 방식으로 사전 및 문법, 다양한 부분적 언어표현을 표상하는 방법으로 이에 대한 상세한 논의는 [19][20]을 참조할 것.

체적으로 DECO 사전보다 수치가 낮게 나타나는 것을 볼 수 있다. 가령 맛집 도메인의 경우, 긍정 어휘 재현율은 각각 47.4%와 69.1%로 차이를 보이며, IT 도메인의 경우는 긍정에서 55.2% vs. 82.3%의 차이를, 부정에서 32.4% vs. 67.7%의 보다 뚜렷한 차이를 보였다.

반면 정확률은 맛집 코퍼스의 부정을 제외하고는 두 경우 서로 유사한 성능을 보여주었는데, 이는 SWN 기반 사전의 경우, 재현율이 매우 저조하기 때문에 제한된 수의 어휘가 중복적으로 실현될 때 정확율의 비중이 높아지기 때문으로 보인다. 예를 들어 IT 코퍼스에서 실현된 긍정 어휘에 대해 SWN 기반 사전을 적용하여 분석한 결과, '좋다'라는 한 단어의 출현 빈도가 51%에 달해 이 경우 상대적으로 재현율 규모가 작은 SWN 기반 사전에서 높은 정확률을 보이게 되는 것이다.

따라서 이러한 왜곡을 보완하기 위해 정확률과 재현율을 통합한 조화 평균(weighted harmonic mean)으로써 F-measure를 사용하였다. 그 결과 SWN 기반 사전은 DECO 기반 사전에 비교해볼 때, 맛집 코퍼스와 IT 코퍼스 긍정-부정을 합한 모두에서 평균 52.7 vs. 72.0으로 나타나 전체적으로 SWN 기반 사전이 DECO 사전에 비해 제한된 성능을 보이고 있음을 확인할 수 있다.

지금까지 살펴본 결과, 영어 SWN을 자동번역해서 감성 사전을 구축하는 것은 우선 해당 감성 어휘 목록이 전혀 제공되지 않는 문제가 있었다. 또한 초기의 어휘목록을 획득하고자 하는 경우 일정 유용한 측면을 가지고 있지만, 이것에 온전히 의존해서 사전을 구성하는 경우, 매우 제한된 재현율의 문제를 극복하기가 어렵다는 점을 환기할 필요가 있다.

5. 결론 및 향후 연구 방향

이상에서 본 연구에서는 영어 SWN에 기반을 두어 한국어 감성사전을 구축하는 과정을 살펴보았다. 영어 SWN의 감성어휘 목록을 구글의 자동 번역 기능을 이용하여 한국어 감성어휘 3,665개를 획득할 수 있었고, 일반적으로 사전을 구축하는 데에 오랜 시간과 비용을 요구한다는 점을 감안할 때, 단기간에 1차적인 감성어휘 목록을 구축하는 유용한 방법이라고 판단할 수 있었다.

그러나 영어 SWN을 기반으로 구축된 한국어 감성사전은 몇 가지 제약을 가지고 있었다. 우선 구축된 3,665개의 단어 중 명사 2,409개가 상대적으로 많아 일반적으로 감성을 표현하는 부류로 알려진 용언(동사, 형용사) 부류가 상대적으로 적게 나타났다는 점이 중요한 한계이다. 그나마 구축된 명사도 대부분 병명이나 인명 등 감성 자체를 나타내는 어휘라고 보기 어려운 경우가 상당수 포함되어 있었다.

또한 이들의 성능도 충분한 신뢰도를 주기 어려운데, 앞서 논의한 바와 같이 맛집 코퍼스와 IT 코퍼스를 대상으로 재현율, 정확률, F-measure를 측정된 결과, 재현율이 현저하게 낮아, 궁극적으로 F-measure에서 낮은 수치의 성능을 보여주었다.

현재 본 연구에서는 SWN 기반 사전과 DECO 기반 사전을 결합하여 한국어 감성사전 SELEX 1.0을 구성하였고,

향후 이를 지속적으로 보완 확장할 예정이다. 현 버전의 재현율, 정확률, F-measure를 측정된 결과가 <표6>에 제시되어 있다.

표6. 코퍼스를 통해 확인한 SELEX 1.0의 재현율, 정확률, F-measure

사전	극성	Recall	Precision	F-measure
맛집	긍정	70.1	85.7	77.1
코퍼스	부정	71.1	46.2	56.1
IT	긍정	83.2	83.6	83.5
코퍼스	부정	69.2	72.9	71.1

위에서 F-measure만을 보았을 때, <표4>와 <표5>의 DECO 기반 사전과 비교하여 수치가 증가한 경우는 맛집 리뷰와 IT 리뷰 코퍼스의 '부정'의 경우밖에 없다. 그것도 각각 0.7, 0.5의 수치만 증가하여 실제로 그 성능 향상의 정도는 매우 미미하며, '긍정'의 경우는 DECO 기반 사전과 비교할 때, 맛집 리뷰는 변동이 없고, IT리뷰의 경우는 오히려 1.2 하락한 것을 볼 수 있다. 이러한 결과는 SWN 기반 사전을 포함한 SELEX 사전이 실제 DECO 사전의 성능과 크게 다르지 않음을, 즉 SWN 기반 사전이 큰 영향을 미치지 않음을 보여준다.

지금까지의 논의를 통해서 볼 때, SWN을 그대로 번역하여 한국어 감성어휘 목록을 획득한다는 것은 사실상 불가능하며, 이러한 번역 오류의 노이즈를 수작업으로 일일이 제거해내는 경우에도 그 재현율의 성능이 현저하게 떨어지는 것을 확인할 수 있다. 본 연구의 논의를 위해 대조군으로 제시한 한국어전자사전 DECO의 성능과 비교하면 그 차이를 뚜렷하게 파악할 수 있다.

이러한 점을 통해서 오피니언 마이닝을 위한 본격적인 한국어 감성사전의 구축을 위해서는 한국어 고유의 언어적 속성을 보다 섬세하게 분석하고 기술하는 연구가 반드시 선행되어야 한다는 것을 확인할 수 있다.

현재 후속 연구로써 SWN과 DECO에 기반을 둔 한국어 감성사전 SELEX의 표제어 및 관련 정보를 계속 확장 보완하고 있으며, 더 나아가서 다양한 실제 코퍼스에 나타난 미분석어 유형의 감성어휘도 지속적으로 추가하고 있다. 이와 더불어 여러 다양한 다국어에 대한 감성사전을 구축하는 방법론을 개발함으로써 다국어 온라인 문서들에 대한 보다 효율적인 오피니언 마이닝 시스템 구축을 위한 다국어 언어자원을 구축하는 연구를 진행할 예정이다.

참고문헌

[1] Dave, K., Lawrence, S. & Pennock, D. M., Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. In Proceedings of International Conference on World Wide Web(WWW-2003), 2003.

[2] Nakagawa, T. & Yi, J., Sentiment Analysis: Capturing Favorability Using Natural Language Processing. In Proceedings of the K-CAP-03, 2nd International Conference on Knowledge Capture,

- 2003.
- [3] Liu, B., *Sentiment Analysis: mining opinions, sentiments, and emotions*. Cambridge Univ Press. 2015.
- [4] Esuli, A & Sebastiani, F., SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining. In *Proceedings of the 5th Conference on Language and Evaluation(LREC 06)*, pp.417-422. 2006.
- [5] Cambria, E., Speer, R., Havasi, C., & A.Hussain. SenticNet: A publicly available semantic resource for opinion mining. In *Proceedings of AAAI CSK*, pp.14-18, 2010
- [6] Strapparava, C., & Valitutti, A., WordNet-Affect: an affective extension of WordNet. In *Proceedings of LREC*, pp.1083-1086, 2004
- [7] Balahur, A. & Turchi, M., Multilingual sentiment analysis using machine translation?. In *Proceedings of the 3rd workshop in computational approaches to subjectivity and sentiment analysis*, pp.52-60, 2012
- [8] 최석재·권오병, “빅데이터 분석을 위한 한국어 SentiWordNet 개발 방안 연구”, *어문논집* 58, pp.127-159, 2014
- [9] 윤애선·권혁철, “감정 온톨로지의 구축을 위한 구성요소 분석”, *인지과학* 21(1), pp.157-175, 2010
- [10] Miller, G., WordNet: An On-Line Lexical Database. *International Journal of lexicography*.
- [11] Baccianella, S., Esuli, A. & Sebastiani, F., SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. 7th *Language Resources and Evaluation Conference (LREC 2010)*, pp.2200-2204, 2010.
- [12] Denecke, K., Using SentiWordNet for Multilingual Sentiment Analysis, In *Proceedings of IEEE 24th International Conference 2008*, pp.507-512, 2008.
- [13] Das, A. & Bandyopadhyay, S., SentiWordNet for Indian Languages. In *Proceedings of the 8th Workshop on Asian Language Resources(ALR)*, pp.56-63, 2010.
- [14] Das, A. & Bandyopadhyay, S., Towards the Global SentiWordNet. In *Proceedings of the 24th Pacific Asia Conference on Language*, pp.799-809, 2011.
- [15] 남지순·어건주·박순강·조동희·김새롬·신동혁·최성화, “다국어 감성주석 코퍼스 구축 방법론 DICORA-TR-2015-05”, 한국외국어대학교, 2015.
- [16] 안애림·심승혜·남지순, “온라인 오피니언 문서 분류를 위한 한국어 형용사 의미 극성 사전”, *한글 및 한국어 정보처리학회*, pp.166-171, 2010.
- [17] Nam, J., *Korean Electronic Dictionary DECO, TR-2015-02 DICORA*. Seoul: Hankuk University of Foreign Studies, 2015.
- [18] Paumier, S.. De la reconnaissance de formes linguistiques a l' analyse syntaxique. Ph.D. dissertation, Univ of PEMLV in France, 2003.
- [19] Gross, M., "The Constructing of Local Grammar" *Finite-State Language Processing*, The MIT Press, pp.329-352, 1997.
- [20] Gross, M., "A bootstrap method for constructing local grammars", *Proceedings of the Symposium on Contemporary Mathematics*, University of Belgrad, pp.229-250, 1999.